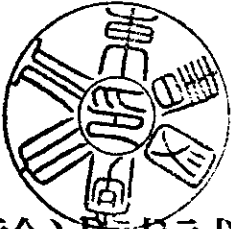




論述式試験の採点システム構築に関する統計科学的研究

著者	柴山 直
URL	http://hdl.handle.net/10097/39616



論述式試験の採点システム構築に関する 統計科学的研究

17300088

平成17年度～平成19年度科学研究費補助金
(基盤研究(B))研究成果報告書

平成20年5月

研究代表者 柴 山 直
東北大学大学院教育学研究科教授

はしがき

ある学習目標が達成できたかできたかどうかを確認する方法として、学習者にその学習成果を実際に遂行してもらい、それを評価するパフォーマンス・アセスメントはさまざまな評価場面において、大きな役割を果たすようになってきている。

また、パフォーマンス・アセスメントは具体的に何ができるようにならなければいけないかが学習者にとって明確になる分、いわゆる教育的意義も大きい。しかし、評定者が人間であるため、その評価にはどこまでも主観的判断がつきまとう。その結果、心理測定論的にみてテスト・スコアの信頼性が客観式テストほどには担保できないのも事実である。

そこで、パフォーマンス・アセスメントの中でも、課題自体の論理構造からいっても、また評価者側の要因からいっても、他のパフォーマンス・アセスメントと比較して、ある程度、要因のコントロールがしやすい小論文の評価に焦点を絞り、まずは、パフォーマンス・アセスメントにおけるスコアリングに共通する基礎的な知見を、統計的に明らかにしようとして始まったのがこの研究である。

さいわい、司法制度の一連の改革の一環として、時を同じくして実施された法科大学院統一適性試験第4部「表現力を測る問題」の答案の提供を、実施団体である財団法人日弁連法務研究財団ならびに社団法人商事法務研究会から、2年間にわたって受けることができた。このなまの答案を、研究チームで組んだ実験デザインに基づいて、熟達した採点チームに採点してもらって得られたのが、ここで報告するデータである。小論文の採点に関する研究の歴史自体はずいぶん古いのが、このように、受験生にとってクリティカルな状況下で得られた答案を採点して得られたデータというのは寡聞にして知らない。もちろん答案の提供に当たっては受験生の同意を得るなど個人情報の保護には十分な対策がとられている。

3年間の研究の結果、当初の目的である、1) 得点調整の統計的手法の検討、2) 主観的判断の信頼性に関する教育測定論的考察、3) AIアプローチによる自動採点システムの開発、のそれぞれについて、一定の成果は得られた。しかしながら、この研究の過程で収集されたデータにはまだまだ多くの情報が埋もれている。研究期間は終了するが、これらを引き続き形にしてパフォーマンス・アセスメント全体に役立つ知見を得るのが次の課題であると認識している。

最後になったが、本研究を遂行するにあたっては、日弁連法務研究財団ならびに商事法務研究会からはひとかたならぬご援助をいただいた。ここに記して感謝の意を表したい。

研究代表者 柴山 直
平成20年5月

研究組織

研究代表者	柴山 直	(東北大学大学院教育学研究科教授)
研究分担者	前田 忠彦	(統計数理研究書データ科学研究系准教授)
	新田 克己	(東京工業大学大学院総合理工学研究科教授)
	町村 泰貴	(北海道大学法科大学院教授)
	野口 裕之	(名古屋大学大学院教育発達科学研究科教授)
	藤本 亮	(静岡大学大学院法務研究科教授)
	醍醐 市朗	(東京大学大学院工学系研究科助手)
	藤田 政博	(政策研究大学院大学大学院政策研究科助教授)

交付決定額（配分額） (金額単位：円)

	直接経費	間接経費	合計
平成 17 年度	7,400 千円		7,400 千円
平成 18 年度	1,600 千円		1,600 千円
平成 19 年度	2,800 千円	840 千円	3,640 千円
総計	11,800 千円	840 千円	12,640 千円

研究発表

(1) 学会発表

- (a) 柴山直・前田忠彦 (2006) 評定者間の差に関する得点調整の試みー小論文データを素材としてー 日本テスト学会第3回大会
- (b) Maeda, Tadahiko (2006) “Assessing Reliability of Essay Scoring in the Writing Section of JLSAT” The Fourth LSAC-JLF Seminar.
- (c) 前田忠彦・柴山直 (2007) 論述式試験の採点者配置デザインと信頼性の評価 日本テスト学会第4回大会

(2) 図 書

- (a) 柴山直・前田忠彦 (2006) 第9章 複数採点者による小論文評価に関する方法論的検討 適性委員会編「法科大学院統一適性試験テクニカルレポート 2006」商事法務 pp.119-131.
- (b) 藤田和恭・新田克己 (2006) 第10章 小論文自動採点の試み 適性委員会編「法科大学院統一適性試験テクニカルレポート 2006」商事法務 pp.132-146.

目次

研究成果の概要	1
第 I 部 成果編	4
1 評定者間の差に関する得点調整の試み	5
2 論述式試験の採点者配置デザインと信頼性の評価	7
3 表現力を測る問題の採点における実践的スコアリング法の検討	13
4 表現力を測る問題の出題方針と採点基準	21
5 複数採点者による小論文評価に関する方法論的検討	39
6 小論文の自動採点の試み	52
7 SVM を利用した小論文の採点支援システム	67
第 II 部 資料編	86
1 平成 17 年度小論文採点実験関係実施資料	87
1.1 実験計画	87
1.2 2005 年「表現力を測る問題」採点会の概要	88
1.3 包括的採点	91
1.3.1 包括的採点の実施方法	91
1.3.2 包括採点者用 論点表	95
1.3.3 包括的採点の例 問題 1	96
1.3.4 包括的採点の例 問題 2	100
1.4 分析的採点	102
1.4.1 分析的採点の実施方法	102
1.4.2 分析的採点基準 問題 1	105
1.4.3 分析的採点シート 問題 1	108
1.4.4 分析的採点基準 問題 2	109
1.4.5 分析的採点シート 問題 2	112
1.5 採点者事後アンケート	113
1.6 研究協力誓約書関係資料 1 (実験 1 & 2)	116

1.7 研究協力誓約書関係資料2(実験3)	117
2 平成18年度小論文採点実験関係実施資料	120
2.1 2006年「表現力を測る問題」採点会の概要	120
2.2 分析的採点	122
2.2.1 分析的採点の実施方法	122
2.2.2 分析的採点基準 問題1	125
2.2.3 分析的採点シート 問題1	127
2.2.4 分析的採点基準 問題2	129
2.2.5 分析的採点シート 問題2	131
2.3 採点者事後アンケート	133
2.4 研究協力誓約書関係資料	135

研究成果の概要

研究の目的

本研究の目的は、論述式試験のための最適な採点システムを設計するという課題を、

- (a) 現実的制約下での最適な採点者配置、
- (b) 適切な採点基準の構成、
- (c) 採点者間の相違を調節するスコアリング法、

の3つの下位課題に分け、統計科学および情報科学的な手法を用いて具体的な解決手続きを提案することになった。

具体的には、(a)においては、必要な精度を確保しつつ効率の良い採点者割り当てを達成する方法の定式化をおこない、(b)においては、包括的評価ならびに分析的評価の比較検討、AI的推論機能を用いた自動採点の可能性を探り、(c)では採点者間の差異を調整する公平なスコアリング法を検討する。あわせて、実データへの適用を通じて提案する採点システムの有効性および、それにより生成されるスコアの信頼性・妥当性・有用性を検証した。

平成 17 年度

(1) 採点デザイン構築法に関する理論的・方法的検討

全ての採点者が全ての採点対象を評価する採点計画ではなく、時間的人的資源の制約下で、一部の評定対象を重複させながら複数の評定者が分担するデザインを基本とした。2種類の互いに異なる内容の記述式問題に対しそれぞれ異なる受験者群からなる答案を無作為抽出により準備した。採点方法は分析的採点と包括的採点の2通りがある。また上とは別に100名の協力者からなる回答者群を準備し、2つの問題ともに回答したものを採点するデザインも設定した。

(2) 採点の実施による採点データの作成

上記の採点デザインに基づき、2005年6月から7月の間に2つの問題に回答した答案を収集した。また同年8月4日から7日にかけて、採点デザインに従い、のべ20名の採点者の協力を得てのべ5160枚におよぶ採点結果を得ることができた。

(3) 採点データの解析

解析には一般化可能性理論からのアプローチと不完全多変量データの分析手法からのアプローチを取った。前者によって採点者間の（採点基準のバラツキに起因する）採点のバラツキを考慮した、採点システム全体の信頼性を推定できた。後者のアプローチにより採点者間の基準の違いを吸収するような換算得点が求められた。

(4) 自動採点システム

電子テキスト化された答案データから AI 的手法による採点システムの開発を行った。成果としては、問題文内容に踏み込んで論理性を判断するシステムの構築、採点基準のコンセプトを定量化する指標の発見、システムによる採点結果と人間の採点結果との高い相関を確認できたことがあげられる。

平成 18 年度

(1) 新たな採点データの作成

平成 18 年 8 月 23 日－25 日、東京：商事法務研究会会議室にて 10 名の採点者のもとに一人 280 枚の小論文答案を採点し新たな採点データを作成した。

(2) 外的評価基準との照合

外的基準として法科大学院統一適性試験の客観的セクションのスコアを用い、採点結果の妥当性を相関分析ならびに一般可能性理論のもとで検討した。

(3) 採点基準間の比較検証

人間が設定した基準による分析的評価結果、包括的評価結果、自動採点に基づく評価結果の相互の一致度および整合性に関する検討を行い、成果の一部は柴山・前田 (2007) で発表した。

(4) 評価システム構築法の整理

平成 19 年度にむけて、実際の試験制度内で実現可能な採点デザインのひな形を提案し、採点基準の望ましい設定の仕方に関する指針を与え、スコアの算出方法とその方法に基づく信頼性評価を整理した。

(5) AI 的アプローチによる自動採点システムの開発

テキストデータからの採点の自動化と採点基準の自動抽出の研究についてはその試作アルゴリズムを具体化し藤田・新田 (2007) でその成果を発表した

(6) 米国における論述型試験の採点・解析法等に関する資料収集

平成 18 年 9 月 18 日・19 日に米国、ペンシルバニア州、ニュータウンにある LSAC を訪問し、論述試験の採点法・解析法、および選抜における利用法に関する資料収集と情報交換を行った。

平成 19 年度

最終年度であり、初期の目的に即して成果をまとめると以下ようになる。

(a) 現実的制約下での最適な採点者配置

必要な制度を確保しつつ効率のよい採点者割り当てを達成する手段として、採点者が担当する対象論文を重複させつつ、一つの論文に対して複数の採点者が採点しながらも、一人の採点者がすべての論文を採点しない、循環型の採点者配置を提案した。

(b) 適切な採点基準の構成

採点者配置デザインと信頼性の評価について、測値をともなうデータに対する一般化可能性理論の適用と ANOVA 的アプローチによる欠測値の補完の両面から信頼性の評価方法を提案した。また、昨年度開発した AI 的推論機能を用いた自動採点システムに、パターン識別手法の一つである SVM(support Vector Machine) の観点から大幅な改良を加えた。その結果、サンプル採点答案数が 200 程度あれば、人間の採点者と同程度の採点が可能であるところまで採点精度を向上させることができた。このことは膨大な採点対象答案がある場合に、このシステムを用いることで採点時間の大幅な短縮が見込めることを意味する。

(c) 採点者間の相違を調節するスコアリング法

上記の最適な採点者配置によって得られたスコア行列を不完全なデータ行列とみなし、採点者間の差異、特に採点者に関する平均と分散を相互に調整する統計的手法をこころみ、目的に即した調整が行われることが確認できた。

第 I 部
成果編

本報告書収録の学術雑誌等発表論文は本ファイルに登録しておりません。なお、このうち東北大学在籍の研究者の論文で、かつ、出版社等から著作権の許諾が得られた論文は、個別に **TOUR** に登録しております。